

Examining How Treatment Fidelity Is Supported, Measured, and Reported in K–3 Reading Intervention Research

Philip Capin¹  · Melodee A. Walker² · Sharon Vaughn¹ · Jeanne Wanzek³

© Springer Science+Business Media, LLC 2017

Abstract Treatment fidelity data (descriptive and statistical) are critical to interpreting and generalizing outcomes of intervention research. Despite recommendations for treatment fidelity reporting from funding agencies and researchers, past syntheses have found treatment fidelity is frequently unreported (e.g., Swanson, *The Journal of Special Education*, 47, 3–13, 2011) in educational interventions and fidelity data are seldom used to analyze its relation to student outcomes (O’Donnell, *Review of Educational Research*, 78(1), 33–84, 2008). The purpose of this synthesis was to examine how treatment fidelity is supported, measured, and reported in reading intervention studies conducted with students at risk or with reading difficulties in grades K–3 from 1995 through 2015. All studies ($k = 175$) were coded to extract and classify information related to (a) the characteristics of the intervention study (e.g., publication year, research design); (b) treatment implementer training and support; (c) treatment fidelity data collection procedures, dimensions (i.e., adherence, quality, receipt, dosage, and differentiation), and levels of treatment fidelity data; and (d) the use of fidelity scores in the analysis of treatment effects. Results indicated that less than half (47%) of the reading intervention studies synthesized reported treatment fidelity data (numeric or narrative). Exploratory analyses showed that several study features were associated with the prevalence of fidelity reporting. Studies reporting treatment fidelity largely measured treatment adherence, and scores were, on average, high. Other dimensions of treatment fidelity (e.g., treatment differentiation), and analyses relating fidelity data to outcomes, were consistently absent from

✉ Philip Capin
pcapin@utexas.edu

¹ Meadows Center for Preventing Educational Risk, University of Texas at Austin, 1912 Speedway, D4900, SZB 228, Austin, TX 78712-1284, USA

² School of Education, Boston University, Boston, MA, USA

³ Department of Special Education, Vanderbilt University, Nashville, TN, USA

the corpus of reading intervention studies reviewed. Recommendations for enhancing how treatment fidelity data in intervention studies are collected and reported are presented.

Keywords Treatment fidelity · Treatment integrity · Fidelity of implementation · Intervention fidelity · Intervention research · Reading intervention

Treatment fidelity has garnered increased attention in educational research and practice over the past three decades. For researchers aiming to develop and test evidence-based practices, treatment fidelity can be an important methodological consideration because it supports the accurate interpretation of treatment effects and can inform considerations about scaling up interventions and generalizing to other populations and settings (Moncher and Prinz 1991; Nelson et al. 2012; O'Donnell 2008). Additionally, treatment fidelity data can be used to examine how treatment fidelity relates to student outcomes (e.g., Al Otaiba and Fuchs 2006; Durlak and DuPre 2008; Kaderavek and Justice 2010). Requirements for treatment fidelity reporting by funding agencies (e.g., National Institutes of Health 2011; U.S. Department of Education 2011), position statements of professional organizations (e.g., National Association of School Psychologists 2007), and a rise in peer-reviewed publications about treatment fidelity (e.g., Fogarty et al. 2014; Gresham 2009; Swanson et al. 2011) reflect the increased emphasis on treatment fidelity in intervention research. For schools, federal legislation (i.e., No Child Left Behind Act 2001; Individuals with Disabilities Education Act 2004; Every Student Succeeds Act 2015) requiring educators to implement evidence-based practices has galvanized interest in fidelity. Although schools do not customarily have treatment groups, schools use fidelity information to monitor whether evidence-based programs are implemented as intended.

Over the past 30 years, researchers have broadened the concept of treatment fidelity to include a set of methodological techniques used to monitor and enhance the reliability and validity of interventions (e.g., Borrelli et al. 2005; Roberts 2016). Treatment fidelity was initially referred to as treatment integrity (or treatment adherence) and defined as the degree to which a treatment is implemented as intended (Yeaton and Sechrest 1981). The construct continues to include treatment integrity; however, Dane and Schneider (1998) expanded treatment fidelity to include other dimensions (i.e., subtypes) of treatment fidelity. Dane and Schneider identified five dimensions of treatment fidelity: adherence (extent to which critical components are implemented as intended; also referred to as treatment integrity), quality (measure of instructional quality separate from adherence to components), exposure (amount of instruction provided; referred to by others as dosage), participant responsiveness (extent to which participants responded to the intended treatment; also referred to as treatment receipt), and program differentiation (extent to which the treatment varies from the comparison condition; Kazdin 1986; O'Donnell 2008).

Despite increased attention to treatment fidelity and agreement among researchers about the consequences of treatment fidelity on experimental validity, reviews have repeatedly shown that a majority of intervention studies do not report treatment fidelity data (Gearing et al. 2011; Gresham et al. 2000; O'Donnell 2008; Sanetti et al. 2013; Swanson et al. 2011). The present study uniquely contributes to the understanding of treatment fidelity by presenting a systematic synthesis of authors' efforts to support and measure treatment fidelity in early reading intervention studies for students with or at risk for reading difficulties. Understanding how treatment fidelity is supported, measured, and reported in reading interventions is important

because treatment fidelity has been found to be a significant predictor of student reading outcomes (O'Donnell 2008), and past studies have shown attempts to measure the effectiveness of reading approaches have been hampered by variation in teacher implementation (e.g., Pressley and Rankin 1994). In the sections to follow, we describe recommendations put forth in psychology, health, and social science that have influenced the conceptualization, measurement, and reporting of treatment fidelity in educational intervention research, as well as the results of past reviews of fidelity reporting in educational intervention research.

Recommendations for Measuring and Reporting Treatment Fidelity

Pioneering works from psychologists and health researchers have influenced recommendations for measuring and reporting treatment fidelity in educational intervention research. Behavioral psychology researchers were the first to mention treatment fidelity (Peterson et al. 1982; Quay 1977), and psychologists Frank Moncher and Ronald Prinz (1991) proposed the first set of specific guidelines for the enhancement of treatment fidelity. These guidelines encouraged researchers to (a) operationally define the treatment, (b) adequately train implementers for treatment delivery using treatment manuals, (c) provide ongoing supervision to treatment implementers, (d) measure adherence to treatment via outside observations, and (e) utilize fidelity data to interpret research findings. Many of these recommendations remain central to more recent investigations of treatment fidelity by health scientists (e.g., Bellg et al. 2004; Borelli et al. 2005; Miller and Rollnick 2014) and educational researchers (e.g., O'Donnell 2008; Swanson et al. 2011).

Building on the original work of Moncher and Prinz (1991), health scientists have advanced the definition, monitoring, and measurement of treatment fidelity over the past 20 years. Key to these advancements was the formation of the National Institutes of Health (NIH) Treatment Fidelity Workgroup in 1999, as part of a consortium of NIH-funded projects focused on improving health-related outcomes. This workgroup put forth several recommendations for researchers related to best practices in treatment fidelity: (1) train and supervise treatment implementers using specific strategies, and assess implementers before treatment delivery to ensure implementers acquire critical skills; (2) measure not only treatment adherence and dosage, but also variation in treatment fidelity among implementers, treatment differentiation, and treatment receipt; (3) collect data on treatment and comparison sessions using audio tapes or observations and conduct exit interviews with comparison group implementers to examine treatment differentiation; and (4) monitor and describe treatment receipt to ensure participants understand and can make use of health-related treatments (Bellg et al. 2004).

Special education researchers have also acknowledged the important role of treatment fidelity and offer recommendations for treatment fidelity reporting in educational research. Recommendations for reporting treatment fidelity were published in *Exceptional Children* for experimental and quasi-experimental group design studies (Gersten et al. 2005) and single case studies (Horner et al. 2005) in special education. Describing the treatment fidelity quality indicators for group studies, Gersten et al. (2005) asserted that assessing and reporting treatment adherence is an “essential” element of high-quality studies (p. 156). These guidelines specified that authors conduct regular instructional observations and use a checklist of the most important features of the intervention to measure adherence. Similarly, Horner et al. (2005, p. 168) called for the “continuous and direct measurement” of treatment fidelity within single case studies. Although not posited as a requirement, Gersten et al. (2005, p.156)

considered assessing and describing quality of implementation (i.e., how well the treatment was implemented) as “desirable” for evidence-based practices high-quality group studies in special education. Adherence and quality of intervention are particularly worthy of measurement in educational research given implementers of educational treatments are often required to enact instructional practices without a scripted protocol (Mowbray et al. 2003) and make countless instructional decisions in short periods of time.

Other educational researchers have suggested that authors investigate the technical properties of fidelity measures and use these measures to examine the relation between treatment fidelity and student outcomes. O’Donnell (2008) describe how fidelity measures were constructed and both assess and report the reliability and validity of the treatment fidelity data collected (Mowbray et al. 2003; O’Donnell 2008). Mowbray et al. (2003) identified several approaches to assessing reliability and validity of fidelity measures, including calculating inter-rater agreement across fidelity raters (reliability) and examining the agreement between two different sources of fidelity information (validity). Additionally, experts have suggested researchers examine the relation between treatment fidelity and student outcomes based on previous research showing treatment fidelity data were associated student outcomes (e.g., National Research Council 2004; O’Donnell 2008).

Although researchers across disciplines define and measure treatment fidelity differently, continuing efforts to develop the concept of treatment fidelity provide educational researchers with an array of methodological considerations to help unpack issues related to treatment implementation. Taken together, the guidelines put forth in psychology, health, and education describe treatment fidelity as a multifaceted concept that includes (a) methods for training, supporting, and assessing treatment implementers; (b) procedures for collecting fidelity data; (c) various dimensions of treatment fidelity (i.e., treatment adherence, dosage, quality, differentiation, and receipt); and (d) techniques for analyzing the relations between fidelity and outcomes.

Treatment Fidelity Reporting in Educational Intervention Research

Reviews of fidelity reporting in intervention research studies began surfacing in the early 1980s (e.g., Lysynchuk et al. 1989) and now cover over 20 topic areas in social, psychological, and behavioral science (Gearing et al. 2011). Several trends are present in the extant reviews of treatment fidelity in educational intervention research. First, school-based interventions under-report treatment fidelity in intervention studies (Gresham et al. 1993; McIntyre et al. 2007; O’Donnell 2008; Swanson et al. 2011; Wheeler et al. 2006). For instance, in the field of learning disabilities, a review of treatment fidelity reporting on academic intervention studies published in select special education journals from 1995 to 1999 found 18% of studies reported treatment fidelity (Gresham et al. 2000). Swanson et al. (2011) reviewed the five highest impact general and special education journals that published intervention studies from 2005 to 2009. They found 47% of studies reported treatment fidelity scores (with similar levels of reporting in both general and special education journals), which demonstrated a notable increase in treatment fidelity reporting relative to the previous learning disabilities synthesis (Gresham et al. 2000).

Another common finding in the intervention literature is that authors infrequently use treatment fidelity data to analyze treatment effects (O’Donnell 2008; Swanson 2011). In education, O’Donnell (2008) located five studies that statistically measured the relation

between treatment fidelity and outcomes among K–12 studies. All five of the studies identified by O'Donnell (2008) reported that higher treatment fidelity was associated with statistically significantly improved student outcomes. Swanson et al. (2011) also investigated how authors incorporated fidelity data into the analysis of intervention effects. Of the 50 studies that reported treatment fidelity data, only two studies used fidelity data to interpret conclusions.

A final trend apparent in past reviews of treatment fidelity is that they did not often collect data for all facets of the construct. For example, most school-based treatment fidelity syntheses limited the construct of fidelity to adherence, omitting other fidelity dimensions such as dosage, quality of implementation, treatment receipt, and treatment differentiation (McIntyre et al. 2007; Wheeler et al. 2006). Swanson et al. (2011) did examine treatment adherence, quality of implementation, and dosage; however, they did not address treatment receipt and treatment differentiation.

Since Swanson et al. (2011) conducted their review of treatment fidelity practices in high impact education journals, several education studies have examined treatment fidelity within intervention studies as a multidimensional construct (Domitrovich et al. 2010; Fogarty et al. 2014; Guo et al. 2016; Hamre et al. 2010; Mendive et al. 2016). For instance, in studying the effects of a multicomponent reading comprehension intervention for middle school students, Fogarty et al. (2014) collected data on the five dimensions of fidelity proposed by Dane and Schneider (1998) and examined the impact of these fidelity dimensions on student outcomes. Confirmatory factor analyses showed all five dimensions of treatment fidelity loaded on a single factor. The authors also found that, on average, the intervention did not create meaningful change on targeted outcomes; however, the intervention led to modest gains when high levels of treatment adherence, quality, differentiation, and receipt were present. The Fogarty et al. study and other recently published studies (e.g., Domitrovich et al. 2010) highlight the important role that multiple dimensions of fidelity can play in studying the effects of reading interventions.

Purpose

Given the expansion of the concept of treatment fidelity and advent of guidelines for fidelity reporting across disciplines, there is a need for a comprehensive examination of treatment fidelity in the reading intervention literature. Past research syntheses of treatment fidelity reporting on school-based interventions confined their investigations of fidelity reporting to a single journal (McIntyre et al. 2007; Sanetti et al. 2013), selected high impact journals (Gresham et al. 2000; Swanson et al. 2011), a specific population of students (Wheeler et al. 2006), or only core curriculum reading programs (O'Donnell 2008). The purpose of this synthesis was to provide a systematic and in-depth account of how and to what extent treatment fidelity was addressed in published reading intervention studies conducted with students in grades K–3 from 1995 to 2015. We identified reading intervention studies in the primary grades because this area of research is well-developed—more early reading intervention research has been conducted and more is known about the effects of these interventions than of interventions for students in later grades (Scammacca et al. 2007). We capitalized on the search results of previous syntheses examining the effects of reading interventions for students in the early grades by Wanzek and Vaughn (2007) and Wanzek et al. (2016) to identify many of the articles for this synthesis. Reviewing studies published over the previous 20 years aligns with What Works Clearinghouse (2014) procedures and provides an extensive,

yet contemporary, set of reading intervention studies ($k = 175$) for examining treatment fidelity. This study provides a unique contribution to the existing literature on treatment fidelity because it is the first such study to examine whether study features (e.g., publication year, research design, journal impact factor) were associated with varying levels of fidelity reporting. Additionally, this study examines how authors trained and supported treatment implementers, multiple dimensions of and the extent to which treatment fidelity was reported, and the ways in which authors used treatment fidelity scores to analyze treatment effects. Thus, we posed the following research questions (RQs):

- RQ1. What proportion of K–3 reading intervention studies published in dissertations or peer-reviewed journals reported treatment fidelity data, and how did the proportion of fidelity reporting vary according to study features (e.g., publication year, research design, journal impact factor)?
- RQ2. What procedures did authors report using to train and support treatment implementers?
- RQ3. What procedures did authors report using to collect observations of fidelity and measure treatment fidelity?
- RQ4. What dimensions of treatment fidelity (i.e., adherence, quality, receipt, dosage, and differentiation) and what levels of treatment fidelity were presented in published intervention studies?
- RQ5. To what extent were treatment fidelity scores used in the analysis of treatment effects?

Method

Identification of Reading Intervention Studies

To review the condition of treatment fidelity reporting in early reading intervention studies, we identified unpublished dissertations and peer-reviewed reading intervention studies published from 1995 to 2015 with students in kindergarten through third grade. We took advantage of a corpus of studies identified in past syntheses that examined the effects of less extensive (15–99 intervention sessions; Wanzek et al. 2016) and extensive (100 or more sessions; Wanzek and Vaughn 2007) reading interventions on reading outcomes and updated the searches conducted for these syntheses to include more recent reading intervention studies. For less extensive reading intervention studies, we analyzed the corpus of studies from the synthesis of treatment effects conducted by Wanzek et al., (2016), which identified articles published from 1995 to 2013. To update this synthesis, we replicated the search and screening process for studies published in 2014 and 2015. For extensive reading intervention studies, Wanzek and Vaughn (2007) conducted a comprehensive search of extensive interventions employing nearly identical search and screening procedures as the less extensive synthesis studies published from 1995 through 2005. We updated the search results from their initial synthesis of the effects of extensive interventions to include studies published from 2006 to 2015.

The previous K–3 syntheses (Wanzek et al. 2016; Wanzek and Vaughn 2007) investigating the effects of reading interventions and our updated search for the present synthesis used systematic procedures for identifying unpublished dissertations and published K–3 reading intervention studies. To maximize the number of potentially pertinent articles identified, Wanzek and colleagues conducted a database search using key population and reading search terms and roots (see Wanzek et al. 2016 for details). Additionally, the authors conducted a

hand search of the last 2 years of ten journals that frequently publish reading intervention studies to confirm the most recent published research was captured. The authors initially screened all abstracts for eligibility. For articles that advanced beyond the initial screening, the full text was reviewed for inclusion in the synthesis based on the following inclusion criteria:

1. Study was a dissertation or published in a peer-reviewed journal in English from 1995 to 2015.
2. Participants were students in grades K–3 (or ages five to nine) and were identified with a learning disability, reading difficulty, or as at risk for reading difficulties (e.g., students with low achievement, low phonemic awareness, language disorders).
3. Studies with additional participants were included when a majority of the sample consisted of targeted students or when data were disaggregated for these students.
4. Intervention targeted early literacy in English (i.e., phonics, fluency, phonemic awareness, vocabulary, reading comprehension, or spelling).
5. Interventions were delivered outside of the general education curriculum and were part of typical school programming (i.e., did not include home, clinic, or camp programs)
6. Experimental, quasi-experimental, and single case designs were included.

The inclusion criteria for the syntheses of the effects of more extensive (Wanzek & Vaughn 2007) and less extensive (Wanzek et al. 2016) reading interventions were identical except the less extensive synthesis included reading interventions with 15–99 sessions, whereas the extensive reading intervention synthesis contained interventions with 100 or more instructional sessions.

Figure 1 provides an overview of the search and screening process for this systematic review of treatment fidelity reporting in early elementary reading intervention studies. Wanzek et al. (2016) identified 128 less extensive reading intervention studies published from 1995 to 2013 for their syntheses of effects. Our search yielded 12 additional articles published in 2014–2015. The original extensive reading intervention synthesis (Wanzek and Vaughn 2007) identified 18 studies published from 1995 to 2005, and the extension search for the present synthesis identified 17 studies. In total, we examined 175 reading intervention studies for treatment fidelity reporting.

Coding Procedures

A three-phase process was employed to code the 175 studies identified. First, we designed a coding document (blinded copy available at website) based on guidelines for fidelity data collection and reporting from psychology (Gresham 2009; Gresham et al. 2000; Moncher and Prinz 1991), health (Bellg et al. 2004; Borelli et al. 2005), and education (Gersten et al. 2005; O'Donnell 2008; Swanson et al. 2011). Second, we employed an iterative improvement process by independently coding a randomly selected set of articles from the corpus, met to discuss discrepancies and areas for improvement, and then refined the code sheet to enhance coding reliability. Third, two authors independently double-coded all studies to extract and classify information related to treatment fidelity and met to reach agreement about all coding discrepancies.

Phase One Code Sheet Development The code sheet developed for this synthesis was informed by past research on treatment fidelity (e.g., Bellg et al. 2004; Moncher and Prinz

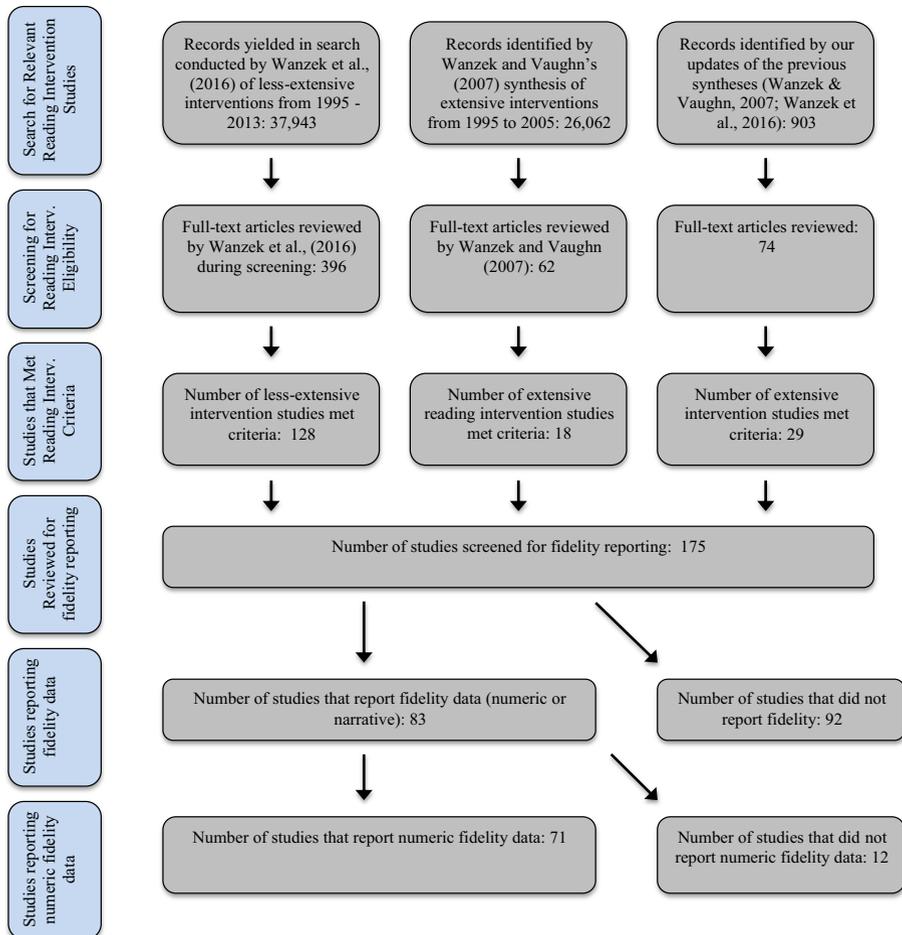


Fig. 1 Identifying reading intervention studies for fidelity synthesis and screening for fidelity data. Studies were excluded during the reading intervention screening phase for not meeting any of the following criteria: (1) a majority of the sample participants were students in grades K through 3 or aged 5 through 9 years or data was disaggregated by grade level for the target grade levels; (2) the reading instruction was provided in an alphabetic language and delivered in a general education classroom; (3) the dependent variable addressed reading performance outcome(s) (i.e., vocabulary, oral reading fluency, comprehension, phonics); (4) the research design was experimental, quasi-experimental, or multiple treatment; and (5) the study was published in English in a peer-reviewed journal from 1995 to 2015. Studies were coded as providing treatment fidelity data when the level of treatment fidelity was numerically or narratively reported

1991) and a code sheet used in a past synthesis of treatment fidelity (Swanson et al. 2011). The final code sheet facilitated the collection of data related to (a) treatment implementer training and support (six items), (b) procedures for collecting treatment fidelity data (six items), (c) treatment fidelity descriptions and scores (eight items), (d) use of fidelity data in analysis (ten items), and (e) general study features (e.g., research design and impact factor of journal; eight items). Professional development sessions that treatment implementers received when beginning (or continuing with) a reading program were considered training for treatment implementers when the training was focused on the treatment. We classified research studies as experiments, quasi-experiments, treatment-comparison, and single case studies. Studies were

classified as quasi-experimental if authors used procedures to establish baseline equivalence (e.g., matching procedures). Studies without randomization or procedures for establishing baseline equivalence were coded as treatment-comparison studies. The 2016 *Journal Citation Reports* (JCR®) from Thomson Reuters were used to identify each study's primary journal category and the journal's impact factor. When a journal's impact factor score was not reported for 2016, we used the most recent impact factor score published by Thomson Reuters.

We coded treatment fidelity data into five dimensions: treatment adherence, quality, dosage, receipt, differentiation, or combined (i.e., when more than one dimension of fidelity was combined into a single score). Data described by authors as an indication of whether the intervention was delivered as intended were labeled adherence data. Information about the level of skill with which a treatment was delivered was categorized as quality data. We classified data related to the number and length of instructional sessions as dosage data. Unlike other dimensions of treatment fidelity data, treatment receipt data were collected based on student behavior (e.g., time on task, student engagement data) rather than implementer behavior. Data comparing treatment and control conditions with one another using common protocols were categorized as treatment differentiation data.

To qualify as a study reporting treatment fidelity, the study was required to explicitly report numeric or narrative information about the level of treatment fidelity of the reading intervention under investigation. Studies that described the fidelity of classroom instruction but did not describe the fidelity of the reading interventions were classified as not providing treatment fidelity data. Articles were also classified as not reporting fidelity data if they reported collecting fidelity data, but provided no description of the fidelity data collected. Articles were not required to use the term *fidelity* nor were articles required to provide numeric fidelity data to be considered studies reporting fidelity data. We classified studies that provided minimal information, such as "the research team noticed high fidelity" as reporting treatment fidelity data.

Phase Two Code Sheet Improvements and Reliability After drafting the initial coding document, the first two authors piloted the code sheet by independently coding three articles and then met to discuss inconsistencies in their coding results and areas for enhancement of the code sheet. Several changes were made to the code sheet to enhance the reliability of data collection. This iterative process of revising the code sheet was repeated twice more with five new randomly selected articles each time. After a third round of revisions to the code sheet, we collected inter-rater reliability data for five new articles. Inter-rater reliability was assessed as the number of items in agreement divided by the total number of items. For each of the sections of the code sheet, an inter-rater reliability of 90% was established as the lowest permissible limit across the items within each section. Actual reliabilities ranged from 91 to 100%, and the overall reliability was 94%. All studies ($n = 175$) were independently double-coded to preserve reliability.

Phase Three Data Collection As articles were double-coded, discrepancies were discussed and decisions were reached by consensus. Code sheet data were then extracted and organized into tables. During the process of organizing information into tables, all numeric treatment fidelity score codes were verified with source articles for a third time.

Data Analysis

To address the first research question, we examined the associations between study features presented as categorical data and treatment fidelity reporting using chi-square statistics. The increased type I error rate associated with testing multiple hypotheses was addressed using the Benjamini and Hochberg (1995) correction, which controls for false discovery rate. Nonetheless, the analyses assessing the association between study features and treatment fidelity reporting should be considered exploratory rather than conclusive due to the dependent nature of these data (discussed in further detail in the Results section). We addressed RQ2 and RQ5 by collecting relevant information from the synthesized studies and presenting this information descriptively. To account for RQ3, we calculated the percentage of articles employing a particular treatment fidelity procedure (e.g., collecting inter-observer agreement data) as the ratio of the number of articles used that procedure to the total number of articles for which the procedure could have been reasonably included. Thus, studies were not included in the denominator when it was not possible for studies to use a particular procedure. For example, we excluded studies that did not use observations to collect fidelity data from the denominator when calculating the percentage of studies that reported inter-observer agreement (refer to Table 3). To attend to RQ4, numeric fidelity scores collected from studies were calculated and reported as percentages to make them comparable across studies (refer to Table 5).

Results

Treatment Fidelity Reporting and Study Features (RQ1)

Of the 175 reading intervention studies coded for this synthesis, 83 studies (47%) reported numeric or narrative treatment fidelity data. Table 1 addresses the first research question by summarizing the key features (e.g., date of publication, research design) of the studies reviewed with and without fidelity data. Below, the results of our exploratory analyses examining the relations between study features and frequency of fidelity reporting are presented.

Publication Year Publication year was statistically significantly related to the occurrence of treatment fidelity reporting, $\chi^2(1, k = 175) = 11.98, p < .01$. As shown in Table 1, the proportion of studies reporting treatment fidelity data increased steadily over time.

Research Design A statistically significant relation between research design and the prevalence treatment fidelity reporting was present, $\chi^2(1, k = 175) = 15.42, p < .01$. Single case was the design most likely to report treatment fidelity data. Fifteen of the 22 (69%) single case studies presented treatment fidelity data. Nearly half of the studies reviewed (48%) were experimental studies and 58% reported treatment fidelity data. This proportion exceeds the share of quasi-experimental (31%) and treatment-comparison (30%) studies reporting fidelity data.

Title of Treatment Implementer The title of the treatment implementer (e.g., teacher or paraprofessional) was also significantly related to the occurrence of fidelity reporting, $\chi^2(1,$

Table 1 Characteristics of the early reading intervention studies reviewed

Variables	All studies ($k = 175$)		Studies reporting fidelity ($k = 83$)		Studies not reporting fidelity ($k = 92$)	
	N	%	N	%	N	%
Publication year						
1995–1999	30	17	8	27	22	73
2000–2004	41	23	15	37	26	63
2005–2009	55	31	30	55	25	45
2010–2015	49	28	30	61	19	39
Primary journal category						
Education	122	70	64	52	58	48
Psychology	36	21	13	36	23	64
Other	17	10	6	35	11	65
Education journal type ($k = 122$)						
General education	55	45	22	40	33	60
Special education	67	55	42	63	25	37
Research design						
Experimental	83	47	47	57	36	43
Quasi-experimental	50	29	16	32	34	68
Treatment-comparison	20	11	5	25	15	75
Single case	22	13	15	68	7	32
Implementers						
Teacher	30	17	11	37	19	63
Paraprofessional	23	13	18	78	5	22
Research team member	89	51	41	46	48	54
Other	33	19	13	39	20	61
Duration of intervention						
Less-extensive (15–99 sessions)	139	79	65	47	74	53
Extensive (100+ sessions)	36	21	18	50	18	50
Student grade level(s)						
K	41	23	20	45	21	55
1	57	33	36	62	21	38
2	19	11	9	47	10	53
3	11	6	5	46	6	54
Multiple grades	47	27	13	28	34	72

Thomson Reuters' *Journal Citation Reports* (JCR®) were used to identify each study's primary journal category (i.e., education, psychology, etc.). Studies were classified as experiments (randomized controlled trials), quasi-experiments (non-randomized but used procedures to establish baseline equivalence), treatment-comparison (non-randomized and no procedures to establish baseline equivalence), and single case studies

$k = 175$) = 11.08, $p < .05$. Studies with paraprofessional implementers reported treatment fidelity more frequently (78%) than studies with research team members (45%), teachers (34%), or other implementers (e.g., community volunteers; 39%). Although there may be a relation between treatment implementer and the frequency of fidelity reporting, there appears to be a threat to the validity of this finding. Patricia Vadasy was the first author on nine of the 22 studies conducted with paraprofessionals, and all of these studies reported treatment fidelity (Vadasy and Sanders 2008a; Vadasy and Sanders 2008b; Vadasy and Sanders 2010; Vadasy and Sanders 2011; Vadasy et al. 2005; Vadasy et al. 2006a; Vadasy et al. 2006b; Vadasy et al. 2007; Vadasy et al. 2002, b). If these studies were removed, the percentage of studies with paraprofessional implementers reporting treatment fidelity (46%) would have approximated the level of treatment fidelity reporting across all studies (47%).

Participant Grade Level Student grade level was also significantly associated with treatment fidelity reporting, $\chi^2(1, k = 175) = 13.07, p < .05$. Studies with participants in grade 1 reported treatment fidelity most frequently (62%), and studies with multiple grade levels reported fidelity data least frequently (28%), whereas the reporting for studies with participants in other grades (ranged from 45 to 47%) were similar to the overall sample rate (47%). Of note, the association between student grade and frequency of fidelity reporting appeared to be confounded by research design. Two thirds of studies (39 of 58; 67%) conducted with first grade students used an experimental design or single case design (which were more likely to report treatment fidelity) whereas just about half of studies (24 of 47; 51%) conducted with students in multiple grade levels used these design types.

Journal Category We also investigated whether there was a relation between the journal category that studies were published in and the probability of treatment fidelity reporting. Journal category (i.e., education, psychology, or other) was not statistically significantly associated with treatment fidelity reporting, $\chi^2(1, k = 175) = 4.09, p = .13$, though descriptive data showed studies with a primary journal category of education had a higher percentage of studies presenting treatment fidelity (52%) than studies published in journals with psychology listed as the primary category (36%) or other journal categories (35%). A closer look at the education journals revealed that the type of education journal (i.e., special or general education) was statistically significantly related to treatment fidelity reporting, $\chi^2(1, k = 122) = 6.23, p < .05$, with journals identified primarily as special education journals reporting treatment fidelity at a higher rate (63%) than general education journals (40%).

Journal Impact Factor Of the 175 studies in the corpus, we were able to locate journal impact factors associated with 160 studies. We were unable to locate journal impact factors for five dissertation studies and ten peer-reviewed journal articles that were published in journals without impact factor information. More missing data were found for studies that did not report treatment fidelity ($k = 10$) than studies that reported fidelity data. Insufficient power precluded a statistical test for differences between the journal impact factor for studies that reported treatment fidelity and those that did not while accounting for the nested nature of the data (intervention studies were nested in authors); therefore, we present descriptive information. The average journal impact factor for studies reporting fidelity appeared slightly higher ($M = 1.82$) than for studies not reporting treatment fidelity ($M = 1.52$) while both sets of studies had similar standard deviations (0.98 and 0.95, respectively) and ranges (0.45–6.62 and 0.42–6.62, respectively). The effect size value (Hedges' $g = 0.32$) suggests that there may be a modest difference between studies with treatment fidelity data reported and those without; however, this apparent difference may be due to chance.

Duration of Intervention The duration of intervention (i.e., 15–99 or 100 or more sessions) was not significantly associated with the frequency of fidelity reporting, $\chi^2(1, k = 175) = 0.12, p > .05$.

Procedures for Enhancing Treatment Fidelity (RQ2)

We investigated the methods that authors reported using to train and support reading intervention implementers (Table 2). A high proportion of studies (77%) that reported treatment

fidelity described the initial training, support provided to treatment implementers after the onset of intervention, or both. In 63 of these 83 studies (76%), authors reported delivering an initial training to implementers before the start of the intervention, and 50 of the studies (60%) reported performing an initial training and supporting implementers after beginning the intervention. In nearly all of these studies, the authors reported that the initial training was provided by the research team or a program trainer for the reading intervention company. For instance, O'Connor et al. (2014) sought to foster fidelity using an initial 4-h training and 2-h bimonthly follow-up meetings led by the research team. Only one early intervention study described providing ongoing support without an initial training (Nunnery et al. 2006). Additionally, 13 of the 83 studies (16%) that reported fidelity data stated that the research team assessed implementer's knowledge of intervention or skills before the start of instruction. Lane et al. (2009), for example, described assessing each implementer's intervention skills through a simulated lesson before implementers began instruction with students.

Many studies included in this synthesis also reported the amount of time spent providing the initial training and ongoing support. Fifty-eight of the 63 studies (70%) that described the initial training stated the number of hours of the initial training meetings for teachers. Authors less frequently (47%) reported the number of hours implementers received training or support after the intervention was underway. Table 2 shows that there was considerable variation in the number of initial training and ongoing support hours that occurred among studies.

Table 2 Treatment implementer training and support to promote treatment fidelity

Variables	Studies reporting treatment fidelity ($k = 83$)	
	<i>N</i>	%
Implementer training		
Training and support	50	60
Training only	13	16
Support only	1	1
Not reported	19	23
Number of training hours before start of intervention ($k = 63$)		
5 h or less	24	38
5.5–10 h	9	14
10.5–15 h	6	10
15.5–20 h	8	12
20.5–25 h	6	10
25 or more hours	5	8
Not reported	5	8
Number of implementer support hours after start of intervention ($k = 51$)		
5 h or less	7	14
5.5 to 10 h	10	20
10.5 to 15 h	2	4
15.5 to 20 h	2	4
20.5 or more hours	3	10
Not reported	27	53
Assessed implementer knowledge or skills before onset of intervention		
Yes	13	16
No	70	84

Procedures for Measuring Treatment Fidelity (RQ3)

To address the third research question, we examined the procedures authors reported using to collect fidelity data (Table 3). A high percentage of studies (90%) included a live, video, or audio observation as the means for collecting fidelity data. Of the studies collecting fidelity data using observations, most of the studies ($k = 63$) reported conducting observations in person (i.e., live); however, some studies reported collecting audio ($k = 8$) and video recording ($k = 2$) of instruction for later coding. All studies that did not report collecting observational data used self-report measures to examine treatment fidelity. For example, Graham et al. (2002) described a self-report fidelity measure in which teachers marked whether the lesson steps were completed after each lesson.

Many of the studies that used observations to collect fidelity data provided information about the frequency and duration of observations and provided inter-observer agreement data to describe the reliability of observational data. Of the studies that reported conducting observations, 54 of the 72 studies (75%) reported the number of observations performed and 52 of the 72 articles (72%) reported the duration of observations (i.e., number of minutes). Across studies, the number of observations ranged from 1 to 35 observations. Of the studies that reported the number of observations, 16 of the 54 studies (30%) reported conducting three or less observations per implementer whereas 29 studies (33%) and 13 studies (17%) reported conducting four to 12 observations and 13 or more observations, respectively. In the studies reviewed, the majority of the interventions lasted 20 or 30 min and the authors reported observing entire intervention sessions when collecting fidelity data. Hence, 37 of the 52 studies

Table 3 Fidelity data collection features

Variables	Studies reporting treatment fidelity ($k = 83$)	
	k	%
Data collection procedure		
Live observation	63	76
Audio observation	8	10
Video observation	2	2
Self-report	8	10
Live observation and self-report	2	2
Inter-observer agreement of observations ($k = 75$)		
90% or more	20	27
80–89%	10	13
80% or less	2	3
Not reported	42	56
Not applicable (only one observer)	1	1
Number of observations per implementer ($k = 75$)		
3 or less	16	21
4–6 observations	8	11
7–9 observations	5	7
10–12	11	15
13 or more	13	17
Not reported	22	29
Duration of observations ($k = 75$)		
15 min or less	5	7
16–30 min	37	49
31 min or more	10	13
Not reported	23	31

(72%) reported observations lasted 16 to 30 min. Inter-observer agreement data were collected in 32 of the 75 studies (43%) that reported performing observations of fidelity. Of the reported inter-observer agreement data, 30 of the 32 studies (94%) reported agreement above 80%, with 20 studies (63%) reporting agreement above 90%.

Dimensions and Levels of Fidelity Data (RQ4)

Dimensions of Treatment Fidelity To address the fourth research question, we examined the various dimensions (e.g., adherence and quality) of treatment fidelity data reported in the corpus. Sixty studies (72%) only reported adherence, one study (1%) reported only quality, eight studies (10%) only reported a combined indicator (e.g., adherence and treatment receipt in a single score), and 14 studies (17%) reported multiple measures of fidelity (e.g., adherence, dosage, treatment receipt indicators reported separately). Of the intervention studies that reported multiple dimensions of fidelity, 13 of the 14 studies included measures of adherence. Thus, nearly 90% of studies (73 of 83) that reported treatment fidelity data included an indicator of treatment adherence. Thirteen of the studies that reported multiple dimensions of treatment fidelity reported quality scores. Although there were no studies that reported only

Table 4 Types of fidelity measures reported and their scores

Variables	Studies reporting treatment fidelity ($k = 83$)	
	k	%
Types of data reported		
Numeric data	71	86
Narrative data only	12	14
Types of fidelity data reported		
Adherence only	60	72
Quality only	1	1
Combined only	8	10
Multiple types	14	17
Types of fidelity data reported in studies with multiple fidelity measures ($k = 14$)		
Adherence	13	93
Quality	13	93
Treatment receipt	3	21
Dosage	2	14
Fidelity scores for studies that reported numeric adherence scores ($k = 65$)		
90 to 100%	50	77
80 to 90%	10	15
70 to 80%	4	6
Less than 70%	1	2
Fidelity scores for studies that reported numeric quality scores ($k = 14$)		
90 to 100%	5	36
80 to 90%	1	7
70 to 80%	5	36
Less than 70%	3	21
Fidelity scores for studies that reported numeric combined scores ($k = 9$)		
90 to 100%	3	33
80 to 90%	5	56
70 to 80%	1	11
Less than 70%	0	0

dosage or treatment receipt as fidelity data, discrete measures of dosage and treatment receipt were included in two and three studies, respectively. Table 4 includes a full account of the types of fidelity data reported.

Levels of Treatment Fidelity Of the 83 studies that reported treatment fidelity data, 71 studies (86%) presented numeric treatment fidelity data while the remaining 12 studies presented qualitative descriptions (see Table 4). The level of treatment fidelity in the studies synthesized was high across the types of data (i.e., numeric or narrative descriptive data) and dimensions of fidelity. All of the studies that provided qualitative descriptions of treatment fidelity described fidelity as high or without substantial deviations from the intended treatment. For instance, Torgesen et al. (2010, p. 46) explained, “Although no formal analysis of fidelity was conducted, information from the videotapes indicated that teacher fidelity to the intervention procedures and materials of both methods was very high throughout the implementation period.”

Examining the numeric treatment fidelity data revealed fidelity was high, especially among studies that measured adherence to the intended treatment. Specifically, 50 of the 65 studies (77%) that reported numeric adherence data asserted adherence at or above 90%, with 32 of the studies (50%) reporting fidelity equal to or exceeding 95%. The majority of remaining studies (10 of 15) that collected adherence data reported adherence data exceeded 80%. Only one study (Denton et al. 2010a) reported adherence fidelity below 70%. Scores for the other dimensions of treatment fidelity were also relatively high. We present fidelity scores for each of the 83 studies along with study feature (i.e., duration of intervention, research design, student grade level, implementer) and fidelity collection information (e.g., method of fidelity collection, number and duration of observations) in Table 5.

Treatment Differentiation Treatment differentiation data were rarely reported in the corpus of studies. Only two studies synthesized directly evaluated the extent to which the treatment condition differed from the comparison condition using a common observation tool (Wolgemuth et al. 2014; Vadasy et al. 2015). For instance, Vadasy et al. (2015) coded for unique and common instructional components when assessing the effects of two approaches to improving vocabulary among kindergarten students to examine the extent to which the treatment conditions vary.

Variation in Treatment Fidelity Among Implementers We also collected data on how frequently authors reported variation in treatment fidelity among treatment implementers. A little less than half of the studies (49%) reported variation in fidelity among treatment implementers. Authors most frequently offered a standard deviation for treatment fidelity mean or minimum and maximum scores as information about the variation in treatment fidelity across implementers (e.g., Vadasy and Sanders 2009; Vernon-Feagans et al. 2012).

Using Treatment Fidelity Data to Examine Treatment Effects (RQ5)

Related to the final research question, authors rarely reported using treatment fidelity in their analysis of treatment effects. Only four of the 83 studies (5%) reported examining the effects of fidelity on treatment effects (Nunnery et al. 2006; Vadasy and Sanders 2009; Vadasy et al. 2015; Wolgemuth et al. 2014). Three of these studies found treatment fidelity was predictive of

Table 5 Study characteristics and fidelity data of studies that reported fidelity data

Studies	Duration of intervention	Research design	Student grade level(s)	Implementer	Numeric data reported	Number of fidelity scores provided	Type of fidelity data reported	Method of Fidelity collection	Inter-observer agreement data (avg.)	Number of observations per implementer (avg.)
Al Otaiba et al. (2005)	LE	QE	K	R	Y	1	A	LO	NR	2
Al Otaiba et al. (2014)	E	E	1	R	Y	1	A	LO	98	3
Alves et al. (2015)	LE	SS	3	R	Y	1	A	LO	NR	6
Berninger et al. (2000)	LE	E	1	R	Y	1	A	AO	NR	10
Berninger et al. (2002)	LE	E	2	R	Y	1	A	AO	NR	10
Berninger et al. (2003)	LE	QE	2	R	Y	2	A	AO	NR	2
Center et al. (1995)	LE	QE	1	O	N	0	C	LO	NR	NR
Denton et al. (2006)	LE	SS	M	R	Y	2	M	LO	≥94	4.5
Denton et al. (2014)	LE	E	1	R	Y	2	M	LO	≥85	NR
Denton et al. (2010a)	E	E	1	T	Y	5	M	LO	80	3
Denton et al. (2011)	LE	E	1	R	Y	2	M	LO	90	2.5
Denton et al. (2010b)	LE	QE	K	O	Y	3	M	LO	NA	3
Dowrick et al. (2006)	LE	SS	1	R	Y	1	A	SR	NA	1
Duff et al. (2014)	LE	E	1	P	Y	2	Q	LO	NR	NR
Ehri et al. (2007)	LE	QE	1	O	Y	1	A	LO	NR	NR
Fien et al. (2015)	E	E	1	P	Y	2	M	LO	NR	NR
Foorman et al. (1997)	E	TC	M	T	N	0	A	SR	NA	NR
Fuchs et al. (2008)	LE	E	1	R	Y	1	A	AO	NR	NR
Gibson et al. (2014)	LE	SS	1	R	Y	1	A	LO	NR	7.5
Gilbert et al. (2013)	LE	E	1	R	Y	1	A	AO	NR	10
Gillon et al. (2007)	LE	QE	K	O	N	0	A	LO	NR	NR
Graham et al. (2002)	LE	E	2	R	Y	1	A	SR	NA	NR
Hagan-burke et al. (2011)	E	E	K	T	Y	2	M	LO	63, 87	3
Jenkins et al. (2004)	E	E	1	P	Y	2	A	LO	96	26
Jones et al. (2009)	LE	SS	3	R	Y	1	A	SR	NA	NR
Kamps and Greenwood (2005)	LE	QE	1	O	Y	1	A	SR	NA	6
Lane et al. (2001)	LE	SS	1	R	Y	2	A	LO	NR	NR
Lane et al. (2001)	LE	SS	1	R	Y	1	A	LO	NR	3
Lane et al. (2009)	LE	E	1	R	Y	1	A	LO	NR	2
Little et al. (2012)	E	E	K	O	Y	1	A	LO	87	3
Lo et al. (2011)	LE	SS	2	R	Y	1	A	LO	NR	NR

Table 5 (continued)

Studies	Duration of intervention	Research design	Student grade level(s)	Implementer	Number of fidelity scores provided	Type of fidelity data reported	Method of Fidelity collection	Inter-observer agreement data (avg.)	Number of observations per implementer (avg.)
Mathes and Babyak (2001)	LE	QE	1	T	1	A	LO	97	3
Mathes et al. (2003)	LE	QE	1	T	1	A	LO	≥88	3
Mathes et al. (2005)	E	E	1	R	4	M	LO	NR	NR
McMaster et al. (2005)	LE	E	1	R	1	C	LO	NR	NR
Morris et al. (2012)	LE	E	M	R	0	A	LO	NR	NR
Musti-Rao and Carledge (2007)	LE	SS	M	O	1	A	LO	100	12.5
Nelson et al. (2005b)	LE	E	K	P	2	A	M	NA	NR
Nelson et al. (2005a)	LE	E	K	R	2	A	M	NA	NR
Niedringhaus (2012)	E	TC	3	T	1	C	LO	NR	3
Noltemeyer et al. (2014)	LE	SS	2	R	1	A	LO	NR	4
Nunery et al. (2006)	E	E	3	T	0	C	LO	NR	NR
O'Connor et al. (2014)	LE	QE	2	P	0	C	LO	NR	NR
O'Connor et al. (2013)	LE	QE	M	T	1	A	LO	NR	5
O'Connor et al. (1995)	LE	E	K	T	1	A	LO	NR	10
O'Connor et al. (1996)	E	QE	K	O	0	A	SR	NA	NR
O'Connor et al. (2010)	LE	E	2	R	0	A	SR	NA	NR
O'Shaughnessy and Swanson (2000)	LE	E	2	P	1	A	LO	NR	15
Oudeans (2003)	LE	E	K	R	1	A	LO	96	54
Percola et al. (2010)	LE	E	1	R	1	A	AO	82 to 96	8
Puhalla (2011)	LE	E	1	R	1	A	LO	90	7
Pullen et al. (2005)	LE	SS	1	R	1	A	AO	100	2
Reisner et al. (2014)	LE	SS	3	T	1	A	SR	NA	NR
Savage et al. (2003)	LE	E	K	P	0	A	LO	NR	NR
Shippen et al. (2008)	LE	QE	M	R	1	A	LO	100	8
Simmons et al. (2007)	E	E	K	O	1	A	LO	85	10
Simmons et al. (2011)	E	E	K	O	2	M	LO	63, 87	NR
Torgesen et al. (2010)	LE	E	1	R	0	A	VO	NR	NR
Ukrainetz et al. (2009)	LE	E	K	R	0	A	VO	NR	NR
Vádasy and Sanders (2008A)	LE	E	M	P	1	A	LO	81	10.33
Vádasy and Sanders (2008B)	LE	QE	K	P	2	M	LO	95	6
Vádasy and Sanders (2009)	LE	E	M	O	2	M	LO	96, 88	18

Table 5 (continued)

Studies	Duration of intervention	Research design	Student grade level(s)	Implementer	Implementer	Numeric data reported	Number of fidelity scores provided	Type of fidelity data reported	Method of Fidelity collection	Inter-observer agreement data (avg.)	Number of observations per implementer (avg.)
Vadásy and Sanders (2010)	LE	E	K	P	P	Y	1	A	LO	97	7
Vadásy et al. (1997)	LE	E	1	C	C	N	0	C	LO	NR	NR
Vadásy et al. (2000)	LE	E	1	C	C	Y	1	A	LO	NR	20
Vadásy and Sanders (2005)	E	TC	1	P	P	Y	1	A	LO	98	20
Vadásy et al. (2015)	LE	E	K	R	R	Y	2	A	LO	96	6
Vadásy et al. (2006b)	LE	E	K	P	P	Y	1	A	LO	90	19
Vadásy et al. (2006a, study 1)	LE	QE	2	P	P	Y	2	A	LO	NR	20
Vadásy et al. (2006a, study 2)	LE	E	M	P	P	Y	2	A	LO	NR	21
Vadásy et al. (2007)	LE	E	M	P	P	Y	1	A	LO	88	16
Vadásy (2011)	LE	E	1	P	P	Y	1	A	LO	97	9.6
Vadásy (2002)	E	TC	M	P	P	Y	1	A	LO	NR	35
Vaughn et al. (2006)	E	E	1	R	R	Y	2	M	LO	95	3
Vernon-Feagan (2012)	LE	QE	M	T	T	Y	1	M	LO	NR	NR
Wang and Algozzine (2008)	E	E	1	P	P	Y	1	A	LO	NR	NR
Wanzek and Vaughn (2008, study 1)	LE	E	1	R	R	Y	1	C	LO	NR	10
Wanzek and Vaughn (2008, study 1)	LE	E	1	R	R	Y	1	C	LO	NR	10
Watson (1997)	LE	SS	M	R	R	Y	1	A	LO	100	NR
Webby et al. (2005)	LE	SS	K	R	R	YB120	1	A	LO	NR	1.5
Webby et al. (2003)	LE	SS	M	T	T	Y	1	A	LO	NR	6
Wolgumuth et al. (2014)	LE	E	K	R	R	Y	2	M	LO	97.98	10

Table 5 (continued)

Studies	Duration of observations (avg.)	Adherence fidelity score (avg.) (%)	Quality fidelity score (avg.) (%)	Treatment receipt fidelity score (avg.) (%)	Dosage fidelity score (avg.) (%)	Combined fidelity score (avg.) (%)	Account for fidelity in analysis?	Report variation in fidelity among implementers?	Report data to support treatment differentiation?
Al Otaiba et al. (2005)	30	98					N	N	N
Al Otaiba et al. (2014)	37.5	89					N	Y	N
Alves et al. (2015)	30	100					N	N	N
Berninger et al. (2000)	20	98					N	Y	N
Berninger et al. (2002)	20	99					N	Y	N
Berninger et al. (2003)	20	95					N	Y	N
Center et al. (1995)	NR						N	N	N
Denton et al. (2006)	NR	96				97	N	N	N
Denton et al. (2014)	45	95	93				N	Y	N
Denton et al. (2010a)	40	69	75	87	83	80	N	Y	N
Denton et al. (2011)	30	73	83				N	Y	N
Denton et al. (2010b)	20		100	98		95	N	N	N
Dowrick et al. (2006)	NR	93					N	N	N
Duff et al. (2014)	NR		74				N	N	N
Ehri et al. (2007)	NR	83					N	Y	N
Fien et al. (2015)	NR	83	78				N	Y	N
Foorman et al. (1997)							N	N	N
Fuchs et al. (2008)	45	98					N	N	N
Gibson et al. (2014)	NR	98					N	N	N
Gilbert et al. (2013)	45	94					N	N	N
Gillon et al. (2007)	NR						N	N	N
Graham et al. (2002)		99					N	N	N
Hagan-burke et al. (2011)	30	77	75				N	Y	N
Jenkins et al. (2004)	30	95					N	N	N
Jones et al. (2009)		100					N	N	N
Kamps and Greenwood (2005)		90					N	N	N
Lane et al. (2001)	30	91					N	Y	N
Lane et al. (2001)	30	98					N	N	N
Lane et al. (2009)	40	93					N	N	N
Little et al. (2012)	NR	80					N	Y	N
Lo et al. (2011)	NR	100					N	N	N

Table 5 (continued)

Studies	Duration of observations (avg.)	Adherence fidelity score (avg.) (%)	Quality fidelity score (avg.) (%)	Treatment receipt fidelity score (avg.) (%)	Dosage fidelity score (avg.) (%)	Combined fidelity score (avg.) (%)	Account for fidelity in analysis?	Report variation in fidelity among implementers?	Report data to support treatment differentiation?
Mathes and Babyak (2001)	30	93					N	Y	N
Mathes et al. (2003)	35	92					N	Y	N
Mathes et al. (2005)	NR	87	87	93		89	N	N	N
McMaster et al. (2005)	NR					89	N	Y	N
Morris et al. (2012)	NR						N	N	N
Musti-Rao and Carledge (2007)	21	100					N	Y	N
Nelson et al. (2005b)	NR	97					N	Y	N
Nelson et al. (2005a)	NR	99					N	Y	N
Niedringhaus (2012)	6					73	N	N	N
Noltemeyer et al. (2014)	NR	100					N	N	N
Nummy et al. (2006)	NR						Y	N	N
O'Connor et al. (2014)	NR						N	Y	N
O'Connor et al. (2013)	25	91					N	N	N
O'Connor et al. (1995)	15	95					N	N	N
O'Connor et al. (1996)							N	N	N
O'Connor et al. (2010)							N	N	N
O'Shaughnessy and Swanson (2000)	30	97					N	N	N
Oudeans (2003)	15	98					N	Y	N
Perricola et al. (2010)	40	90					N	Y	N
Puhalla (2011)	20	96					N	Y	N
Pullen et al. (2005)	30	100					N	N	N
Reisner et al. (2014)		92					N	N	N
Savage et al. (2003)	NR						N	N	N
Shippen et al. (2008)	20	95					N	Y	N
Simmons et al. (2007)	30	90					N	N	N
Simmons et al. (2011)	3	76	75				N	Y	N
Torgesen et al. (2010)	NR						N	N	N
Ukrainetz et al. (2009)	NR						N	N	N
Vadasy and Sanders (2008A)	30	92					N	N	N
Vadasy and Sanders (2008B)	30	95	95				N	Y	N
Vadasy and Sanders (2009)	30	90	92				N	Y	N

Table 5 (continued)

Studies	Duration of observations (avg.)	Adherence fidelity score (avg.) (%)	Quality fidelity score (avg.) (%)	Treatment receipt fidelity score (avg.) (%)	Dosage fidelity score (avg.) (%)	Combined fidelity score (avg.) (%)	Account for fidelity in analysis?	Report variation in fidelity among implementers?	Report data to support treatment differentiation?
Vadasy and Sanders (2010)	30	88					N	Y	N
Vadasy et al. (1997)	NR						Y	N	N
Vadasy et al. (2000)	30					89	N	N	N
Vadasy and Sanders (2005)	30	96					N	N	N
Vadasy et al. (2015)	30	95					Y	N	Y
Vadasy et al. (2006b)	30	91					N	Y	N
Vadasy et al. (2006a, study 1)	30	97					N	N	N
Vadasy et al. (2006a, study 2)	30	95					N	Y	N
Vadasy et al. (2007)	30	95					N	Y	N
Vadasy (2011)	30	90					N	Y	N
Vadasy (2002)	30	92					N	N	N
Vaughn et al. (2006)	40	86	96				N	Y	N
Vernon-Feagan (2012)	15	54	54	58			N	Y	N
Wang and Algozzine (2008)	NR	93				91	N	Y	N
Wanzek and Vaughn (2008, study 1)	30						N	Y	N
Wanzek and Vaughn (2008, study 1)	30					85	N	Y	N
Watson (1997)	NR	100					N	Y	N
Webby et al. (2005)	20	100					N	Y	N
Webby et al. (2003)	30	97					N	Y	N
Wolgemuth et al. (2014)	30	88	51				Y	Y	Y
Avg:		93	81	93	71	88			

LE less-extensive (i.e., 15–99 sessions), E extensive (i.e., 100 or more sessions), E experimental, QE quasi-experimental, TC treatment-comparison, SS single subject, K kindergarten, R researcher, O other, T teacher, P paraprofessional, M multiple, Y yes, N no, A adherence, C combined, M multiple, Q quality, LO live observation, AO audio observation, SR self-report, VO video observation, NR not reported, Avg. average

learning gains for all students (Vadasy and Sanders 2009; Vadasy et al. 2015) or a subgroup of students (Nunnery et al. 2006) on a majority of outcome measures. Vadasy and Sanders (2009) assessed treatment adherence and quality and found both dimensions accounted for differences in reading gains for students in grades 2 and 3 with reading difficulties. Similarly, in a study with kindergarten students at risk for reading problems, Vadasy et al. (2015) observed that treatment effects were moderated by fidelity on a majority of measures at posttest, with greater adherence to treatment associated with improved outcomes. In an effectiveness trial investigating the impact of Accelerated Reader and Renaissance Reading, Nunnery et al. (2006) found implementation fidelity (modeled as a single factor with several components including dosage and quality) did not account for variance in student outcomes for the full sample. However, higher levels of treatment fidelity were associated with improved outcomes for students with disabilities. Wolgemuth et al. (2014) found intervention quality and adherence only predicted student outcomes on two of the four early reading measures used. The authors note that restriction of range may have played a role in why instructional quality and adherence did not predict student outcomes, as implementation was, on average, high with little variability.

Discussion

The purpose of this synthesis was to describe fidelity reporting in early reading intervention studies published from 1995 to 2015. We were interested in (a) the proportion of studies that reported treatment fidelity and whether this proportion varied according to study features (RQ1), (b) the methods that authors used to support high levels of treatment fidelity (RQ2), (c) procedures used to assess treatment fidelity (RQ3), (d) the dimensions and levels of treatment fidelity reported (RQ4), and (e) to what extent treatment fidelity scores were used in the analysis of treatment effects (RQ5). In examining how K–3 reading intervention studies addressed treatment fidelity, we coded 175 studies that met the inclusion criteria and employed a systematic process for obtaining and classifying relevant information.

Treatment Fidelity Reporting and Study Features

Overall Rate of Fidelity Reporting Our first research question addressed the rate of treatment fidelity data reporting. In agreement with previous research findings (Gresham et al. 2000; McIntyre et al. 2007; O'Donnell 2008; Swanson et al. 2011; Wheeler et al. 2006), we found that authors reported treatment fidelity data in less than half of the intervention studies under review. Of the 175 studies reviewed, 83 studies (47%) reported treatment fidelity data. This percentage matched the proportion of studies reporting treatment fidelity scores in Swanson et al. (2011) review of intervention studies published in high impact education journals from 2005 to 2009 and exceeded previous reviews of fidelity reporting in school-based intervention studies (Gresham et al. 2000; McIntyre et al. 2007; Wheeler et al. 2006).

Association Between Study Features and Fidelity Reporting Our exploratory analyses suggested that several study features are associated with fidelity reporting. For one, publication year was associated with fidelity reporting, with results indicating fidelity reporting increased over time. This finding was expected given the quality of educational research methodology

has improved over time (Scammacca et al. 2016; Vaughn and Swanson 2015). It also aligns with the previous research showing that syntheses conducted more recently (Swanson et al. 2011; Wheeler et al. 2006) found higher numbers of treatment fidelity reported than earlier syntheses (Gresham et al. 1993; Gresham et al. 2000).

Study design was also significantly associated with fidelity reporting. Consistent with previous research (Swanson et al. 2011), single case studies (68%) were the most likely to report treatment fidelity data, followed by experimental studies (57%), and then quasi-experimental and treatment-comparison studies (32 and 25%, respectively). The rate of fidelity reporting among experimental studies exceeded that of quasi-experimental studies and other treatment-comparison studies. Considering treatment fidelity reporting is a quality indicator (Gersten et al. 2005), and experimental studies are, on average, associated with larger sample sizes, greater use of standardized measures, and other features of high-quality studies (Cheung and Slavin 2016) that may help to explain why experimental studies included treatment fidelity at a higher rate than quasi-experimental and other treatment-comparison studies. One reason for differences in the reporting of treatment fidelity between single case studies and experimental studies—both design types are considered high-quality—may be related to the category of education journal (e.g., special education, psychology). All of the single case studies were published in special education journals whereas the experimental studies were published in special and general education, as well as psychology journals. Our findings indicated studies published in special education journals were more likely to report treatment fidelity than studies published in general education journals.

Although we were unable to statistically test for a difference in the journal impact factor for studies with and without treatment fidelity data, the modest effect size ($g = 0.32$) difference in the average journal impact factors favoring studies reporting treatment fidelity suggests that future research may be warranted. Research in medical science has found that higher impact journals are associated with more methodologically sound studies (Gluud et al. 2005) and are more likely to include studies with methodological practices intended to safeguard against bias (Bala et al. 2013). Thus, we hypothesized that the studies reporting treatment fidelity would be, on average, published in higher impact journals because reporting treatment fidelity is thought of as an indicator of study quality (Gersten et al. 2005; Horner et al. 2005) and higher impact journals are associated with methodologically higher quality studies. Taken together, the results of these exploratory analyses suggest that some study features (i.e., publication year, research design, and publication year) are related to treatment fidelity reporting.

Implementer Training and Support

Our second research question focused on the training and support authors reported providing to enhance treatment fidelity. This is the first synthesis of treatment fidelity to examine the methods that authors use to support high levels of treatment fidelity. Of the studies reporting treatment fidelity data, a high number of studies (77%) included descriptions of the training provided to treatment implementers. Moreover, among studies that reported providing initial training to implementers, a very high number of studies stated the number of training hours (92%). This level of detail facilitates future study replications by researchers who did not conduct the original investigation (Coyne et al. 2016), which is important because past syntheses have found replications are very limited in the educational research (Lemons et al. 2016; Makel et al. 2016). Although a high number of authors in the corpus of studies detailed

initial training efforts, few studies (16%) reported assessing implementer knowledge or skills before the start of the intervention as a method for enhancing implementer treatment fidelity, as suggested by Bellg et al. (2004). Moreover, less than half of the studies reporting treatment fidelity reported the number of hours implementers received training or support after the intervention was underway. These two areas for improvement are of practical importance as they are relevant to increasing fidelity and improving the quality of reading interventions.

Treatment Fidelity Data Collection Procedures

Of the 83 studies reporting treatment fidelity, authors in all studies described how fidelity data were collected. In line with recommendations (Bellg et al. 2004; Gersten et al. 2005; McIntyre et al. 2007; Moncher and Prinz 1991), authors in 90% of studies employed live, video, or audio observations to collect treatment fidelity data. Past researchers have suggested that fidelity observations occur frequently and over the course of the entire intervention (e.g., Gersten et al. 2005; Horner et al. 2005). Although results show the number and duration of observations conducted to collect fidelity data were reported in most studies, authors did not consistently report whether observations occurred over the course of the entire intervention implementation, making it difficult to determine whether current treatment fidelity data collection procedures met extant recommendations. One may question whether the number of observations in the reviewed studies met the number required to obtain a reliable estimate of treatment fidelity; however, it is difficult to draw conclusions about the reliability of the observational data collected given the limited information presented in intervention studies. One way authors provided reliability information was through inter-observer agreement data. However, in contrast with the recommendations by Gersten et al. (2005) and Horner et al. (2005), a majority of studies (56%) that used observational data failed to report the level of inter-observer agreement. Our findings suggest that most studies collect multiple direct measurements of teacher treatment fidelity, yet additional information about the reliability of treatment fidelity measures is needed.

Treatment Fidelity Dimensions and Scores

In studies reporting treatment fidelity data, adherence data and quality of implementation were the most frequently reported dimensions. The concept of treatment fidelity was once limited to treatment adherence (or integrity; Yeaton and Sechrest 1981), and adherence was the only dimension of treatment fidelity identified as an “essential quality indicator” for group research in special education by Gersten et al. (2005, p. 152). Therefore, it reasons that treatment adherence would be the most common measure of treatment fidelity. Consistent with expert recommendations that the quality of implementation is another dimension identified in the quality standards for group research, albeit as a “desirable quality indicator” (Gersten et al. 2005, p. 152), is the finding of this synthesis that treatment quality is the second most commonly reported dimension of treatment fidelity. Authors less frequently reported other dimensions of treatment fidelity (i.e., treatment dosage, differentiation, and receipt). This may be a product of the lack of emphasis on these dimensions of fidelity in the quality indicators for special education research.

To our knowledge, none of the past reviews of treatment fidelity in school-based intervention studies presented scores for the level of treatment fidelity for all reviewed studies. The results of this systematic review suggest that the treatment fidelity scores reported in reading

intervention studies is typically quite high. The average adherence score across studies was 93%. Scores on treatment quality and dosage measures were lower (81 and 71%, respectively); however, only a few studies reported scores for these dimensions.

Using Treatment Fidelity Data to Examine Treatment Effects

Only four studies examined the effects of fidelity on treatment effects using statistical methods (Nunnery et al. 2006; Vadasy et al. 2015; Vadasy and Sanders 2009; Wolgemuth et al. 2014). Of note, three of these four studies found that treatment fidelity was predictive of learning gains for all students (Vadasy and Sanders 2009; Vadasy et al. 2015) or a subgroup of students (Nunnery et al. 2006). Researchers have suggested that using fidelity data to analytically examine treatment effects provides an important pathway for understanding the hypothesized relations between fidelity and student outcomes (Mowbray et al. 2003; Zvoch 2012). One reason why only a few of the reviewed studies examined the relation between treatment fidelity and student outcomes may relate to the type of experimental trials (e.g., efficacy or effectiveness) being conducted. Researchers conducting efficacy trials typically aim to test an instructional program under ideal conditions, which includes maximizing the level of treatment fidelity and minimizing variation. Thus, these researchers may not hypothesize that treatment fidelity would be related to student outcomes. On the other hand, effectiveness trials may be particularly interested in the relation between treatment fidelity and student outcomes, as these studies aim to test the effects of an intervention under less controlled conditions. Although we do not know how many of the studies reviewed were intended to serve as effectiveness studies, it may be worth noting that over 40% of the treatment implementers were school personnel. This figure, along with previous research (O'Donnell 2008) and the present synthesis findings, reveals that treatment fidelity is associated with student outcomes. This suggests that it may be prudent to include an investigation of the role of treatment fidelity on student outcomes in future reading intervention studies, especially in effectiveness trials.

Limitations and Future Research

The findings should be considered in light of several caveats. For one, although the systematic literature searches intended to be all-inclusive, relevant K–3 reading intervention studies may have been unidentified. Another limitation relates to the collection of treatment fidelity information. Studies in the current corpus typically described treatment fidelity in the methods section of intervention studies; however, this information also appeared in other study sections, making coding difficult. We attempted to enhance the data collection process in four ways: (a) we developed a code sheet through an iterative improvement process before coding data were collected, (b) we established 90% reliability before coding, (c) we independently double-coded all articles and met to discuss discrepancies, and (d) the first author checked all numeric data against the source articles as a final check. Developing clear standards for treatment fidelity reporting would facilitate the future aggregations of fidelity data.

This synthesis may have also been impacted by publication bias favoring the reporting of treatment fidelity data when it was high and the omission of fidelity scores when they were low. Studies may have investigated the effects of treatment fidelity reporting in their analysis of effects but omitted these results when they were non-significant. Past researchers have posited that some intervention articles may not report treatment fidelity data due to journal page limits or a low priority by journal editors on such information (Moncher and Prinz 1991;

Pereplechikova et al. 2007). Thus, the present corpus of studies may not fully reflect the amount of treatment fidelity data collected in reading intervention studies. Additionally, although a substantial amount of intervention research has been conducted in Grades K–3, these findings may not generalize to studies conducted with students in later grades. Future research may aim to investigate treatment fidelity in later grades to determine if these findings generalize.

We also note that it is important to consider the findings related to the associations between study features and fidelity reporting as preliminary. Our results suggest that some study features are associated with the incidence of fidelity reporting; however, we identified instances where these relations appeared to be confounded by third variables. For instance, we found the relation between implementer and frequency of fidelity reporting appeared to be confounded by the high number of studies conducted by a single researcher who consistently reported treatment fidelity data. To generate more conclusive findings about the relations between study features and the frequency of treatment fidelity reporting, future research should investigate these relations in a larger corpus of studies using more advanced methods that account for nested data.

Finally, future researchers should consider how intervention type (e.g., reading comprehension, vocabulary) is related to treatment fidelity. This research could address, for instance, whether higher or lower levels of treatment fidelity are found for certain intervention types. Researchers could also investigate whether treatment fidelity is more strongly associated with outcomes for certain types of intervention.

Implications

Despite progress in the frequency of treatment fidelity reporting over time, the moderate level of reporting in the most recent years reveals that it is not standard practice in the reading intervention literature nor is it customary for researchers to report multiple dimensions of treatment fidelity. Researchers have long recognized that treatment fidelity data are vital to experimental validity (e.g., Shadish et al. 2002); yet, as several researchers have noted (Dane and Schneider 1998; Dusenbery et al. 2003), it appears that many researchers conducting intervention research assume rather than evaluate treatment fidelity. Assuming perfect treatment fidelity may lead to inaccurate interpretations of study findings. For instance, a treatment found to have a null effect on student outcomes relative to a comparison condition may be considered ineffective; however, it may have been the case that the treatment instruction was implemented with low adherence or quality (e.g., O'Donnell 2007). The study would provide clear information that the treatment was not effective relative to the comparison condition when implemented with the low fidelity, but it is unknown how the treatment would have fared given higher instructional quality.

Notwithstanding the considerable research in education and other fields illustrating the complex, multidimensional nature of treatment fidelity (e.g., Bellg et al. 2004; Gresham 2009; Sanetti and Kratochwill 2009), the studies reporting treatment fidelity data typically restricted their examinations of treatment fidelity to measuring treatment adherence. Providing information about instructional quality, treatment receipt (e.g., student engagement), dosage, and treatment differentiation provide a deeper understanding of how the treatment was implemented and received, which can benefit future practitioners focused on implementing the treatment and future researchers aiming to replicate or make alterations the treatment. Denton et al. (2010a) provide an illustrative example of a study that presents a detailed and

multidimensional report of fidelity. The authors were expansive in their description of implementer training and coaching support, describing the number of training hours and focus of the training, as well as multiple dimensions of treatment fidelity, namely treatment adherence, quality, receipt, dosage, and differentiation. To provide treatment differentiation data, Denton and colleagues used the Instructional Content Emphasis-Revised (ICE-R; Edmonds and Briggs 2003) observation tool with both treatment and comparison implementers. Although the authors did not report using fidelity scores to examine treatment effects, this study provides a glimpse of what multidimensional fidelity reporting may include.

Among studies reporting treatment fidelity, authors regularly described methods for collecting treatment fidelity data. Observations were the most common data collection method used in this corpus of studies; however, a majority of studies failed to provide reliability (e.g., inter-observer agreement) or validity (e.g., convergent validity) data related to these observations. Research on the use of classroom observation tools to evaluate teacher quality indicates producing dependable instructional observation tools is a tall task (e.g., Hill et al. 2012). Future studies might examine the psychometric properties of treatment fidelity measures (Sheridan et al. 2009). Studies examining the consistency and sensitivity of treatment fidelity items and their utility in predicting student outcomes may produce more reliable and valid treatment fidelity measures. Improved measures may also result in greater sensitivity to differences in treatment fidelity and, consequently, variation in treatment fidelity scores. Moreover, future intervention studies that incorporate improved treatment fidelity measures may be more apt to accurately measure the relation between treatment fidelity and student outcomes.

For a field that places a high priority on conducting rigorous research, we believe that it is problematic that fewer than half of the studies reviewed included treatment fidelity data. In 2009, Sanetti and Kratochwill put forth several policy recommendations to journal editors and granting agencies about ways to enhance the quantity and quality of treatment fidelity reporting. The findings of this synthesis underscore the continued relevance of these policy recommendations and the need for clear and comprehensive standards for treatment fidelity reporting. Ultimately, thorough consideration and reporting of treatment fidelity is critical to the internal and external validity of reading intervention studies, which in time will help us to better identify evidence-based practices and the conditions under which these practices enhance student outcomes.

Acknowledgements The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institutes of Health, or the Institute of Education Sciences.

Funding Information This research was supported by grants 2P50 HD052117-11 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and R305A150407 from the Institute of Education Sciences, US Department of Education, to The University of Texas.

References

Studies preceded by an asterisk were included in synthesis.

- Al Otaiba, S., & Fuchs, D. (2006). Who Are the Young Children for Whom Best Practices in Reading Are Ineffective? An Experimental and Longitudinal Study. *Journal of Learning Disabilities*, 39(5), 414–431.
- *Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children*, 81(1), 11–27.

- *Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: how much is enough? *Exceptionality*, *13*, 195–208.
- *Alves, K. D., Kennedy, M. J., Brown, T. S., & Solis, M. (2015). Story grammar instruction with third and fifth grade students with learning disabilities and other struggling readers. *Learning Disabilities: A Contemporary Journal*, *13*(1), 73–93.
- Bala, M. M., Akl, E. A., Sun, X., Bassler, D., Mertz, D., Mejza, F., et al. (2013). Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *Journal of Clinical Epidemiology*, *66*(3), 286–295.
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*(5), 443.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.
- *Berninger, V. W., Vaughan, K., Abbott, R. D., Brooks, A., Begayis, K., Curtin, G., et al. (2000). Language-based spelling instruction: Teaching children to make multiple connections between spoken and written words. *Learning Disability Quarterly*, *23*(2), 117–135.
- *Berninger, V. W., Abbott, R. D., Vermeulen, K., Ogier, S., Brooksher, R., Zook, D., & Lemos, Z. (2002). Comparison of faster and slower responders to early intervention in reading: Differentiating features of their language profiles. *Learning Disability Quarterly*, *25*(1), 59–76.
- *Berninger, V. W., Vermeulen, K., Abbott, R. D., McCutchen, D., Cotton, S., Cude, J., et al. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving second-grade readers. *Language, Speech, and Hearing Services in Schools*, *34*(2), 101–116.
- Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., et al. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, *73*(5), 852.
- *Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of reading recovery. *Reading Research Quarterly*, 240–263.
- Cheung, A. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292.
- Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education*, *37*(4), 244–253.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.
- *Denton, C. A., Cirino, P. T., Barth, A. E., Romain, M., Vaughn, S., Wexler, J., et al. (2011). An experimental study of scheduling and duration of “Less-extensive” first-grade reading intervention. *Journal of Research on Educational Effectiveness*, *4*(3), 208–230.
- *Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities*, *39*(5), 447–466.
- *Denton, C. A., Fletcher, J. M., Taylor, W. P., Barth, A. E., & Vaughn, S. (2014). An Experimental Evaluation of Guided Reading and Explicit Interventions for Primary-Grade Students At-Risk for Reading Difficulties. *Journal of Research on Educational Effectiveness*, *7*(3), 268–293.
- *Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E. A., Kethley, C., Kurz, T. B., & Shih, M. (2010a). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, *76*(4), 394–416.
- *Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010b). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *The Elementary School Journal*, *110*(4), 423–439.
- Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*(3), 284–298.
- *Dowrick, P. W., Kim-Rupnow, W. S., & Power, T. J. (2006). Video feedforward for reading. *The Journal of Special Education*, *39*(4), 194–207.
- *Duff, F. J., Hulme, C., Grainger, K., Hardwick, S. J., Miles, J. N., & Snowling, M. J. (2014). Reading and language intervention for children at risk of dyslexia: a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *55*(11), 1234–1243.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350.

- Dusenbery, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256.
- Edmonds, M., & Briggs, K. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn & K. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31–52). Baltimore, MD: Paul H. Brookes.
- *Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: an effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal, 44*, 414–448.
- Every Student Succeeds Act of 2015, Pub. L. No. 114-95.
- Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., & Chaparro, E. (2015). An examination of the efficacy of a multitiered intervention on early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities, 48*(6), 602–621.
- Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., et al. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: a focus on treatment fidelity. *Educational Psychology Review, 26*(3), 425–449.
- *Foorman, B. R., Francis, D. J., Winkates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1*(3), 255–276.
- *Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making “secondary intervention” work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing, 21*(4), 413–436.
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79–88.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children, 71*(2), 149–164.
- *Gibson Jr., L., Cartledge, G., Keyes, S. E., & Yawn, C. D. (2014). The effects of a supplementary computerized fluency intervention on the generalization of the oral reading fluency and comprehension of first-grade students. *Education and Treatment of Children, 37*(1), 25–51.
- *Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly, 48*(2), 135–154.
- Gillon, G. T., Moran, C. A., Hamilton, E., Zens, N., Bayne, G., & Smith, D. (2007). Phonological awareness treatment effects for children from low socioeconomic backgrounds. *Asia Pacific Journal of Speech, Language and Hearing, 10*(2), 123–140.
- Gluud, L. L., Sørensen, T. I., Gotzsche, P. C., & Gluud, C. (2005). The journal impact factor as a predictor of trial quality and outcomes: cohort study of hepatobiliary randomized clinical trials. *The American Journal of Gastroenterology, 100*(11), 2431–2435.
- *Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology, 94*, 669–686.
- Gresham, F. M. (2009). Evolution of the Treatment Integrity Concept: Current Status and Future Directions. *School Psychology Review, 38*(4), 533–540.
- Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22*, 254–273.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research. Practice, 15*(4), 198–205.
- Guo, Y., Dynia, J. M., Logan, J. A., Justice, L. M., Breit-Smith, A., & Kaderavek, J. N. (2016). Fidelity of implementation for an early-literacy intervention: Dimensionality and contribution to children’s intervention outcomes. *Early Childhood Research Quarterly, 37*, 165–174.
- *Hagan-Burke, S., Kwok, O. M., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education, 45*(3), 131–148.
- Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers’ language and literacy growth. *Early Childhood Research Quarterly, 25*(3), 329–347.

- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56–64.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165–179.
- Individuals with Disabilities Education Act of 2004, P.L. 108-446.
- *Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading*, 8(1), 53–85.
- *Jones, K. M., Wickstrom, K. F., Noltemeyer, A. L., Brown, S. M., Schuka, J. R., & Therrien, W. J. (2009). An experimental analysis of reading fluency. *Journal of Behavioral Education*, 18(1), 35–55.
- Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology*, 19(4), 366–379.
- *Kamps, D. M., & Greenwood, C. R. (2005). Formulating secondary-level reading interventions. *Journal of Learning Disabilities*, 38(6), 500–509.
- Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, 54, 95–105.
- *Lane, H. B., Pullen, P. C., Hudson, R. F., & Konold, T. R. (2009). Identifying essential instructional components of literacy tutoring for struggling beginning readers. *Literacy Research and Instruction*, 48, 277–297.
- *Lane, K. L., O'Shaughnessy, T. E., Lambros, K. M., Gresham, F. M., & Beebe-Frankenberger, M. E. (2001). The efficacy of phonological awareness training with first-grade students who have behavior problems and reading difficulties. *Journal of Emotional and Behavioral Disorders*, 9(4), 219–231.
- Lemons, C. J., King, S. A., Davidson, K. A., Berryessa, T. L., Gajjar, S. A., & Sacks, L. H. (2016). An inadvertent concurrent replication: Same roadmap, different journey. *Remedial and Special Education*, 37(4), 213–222.
- *Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O. M., Hagan-Burke, S., Simmons, L. E., Fogarty, M., Oslund, E., & Coyne, M. D. (2012). A comparison of responsive interventions on kindergarteners' early reading achievement. *Learning Disabilities Research and Practice*, 27(4), 189–202.
- *Lo, Y. Y., Cooke, N. L., & Starling, A. L. P. (2011). Using a repeated reading program to improve generalization of oral reading fluency. *Education and Treatment of Children*, 34(1), 115–140.
- Lysynchuk, L. M., Pressley, M., d'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. *Reading Research Quarterly*, 458–470.
- Makel, M. C., Plucker, J. A., Freeman, J., Lombardi, A., Simonsen, B., & Coyne, M. (2016). Replication of special education research: Necessary but far too rare. *Remedial and Special Education*, 37(4), 205–212.
- *Mathes, P. G., & Babyak, A. E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice*, 16, 28–44.
- *Mathes, R. G., Torgesen, J. K., Clancy-Menchetti, J., Santi, K., Nicholas, K. R., Robinson, C., & Grek, M. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *The Elementary School Journal*, 103(5), 459–479.
- *Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, 40, 148–182.
- *McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: an experimental field trial of identification and intervention methods. *Exceptional Children*, 71, 445–463.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659–672.
- Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology*, 108(1), 130.
- Miller, W. R., & Rollnick, S. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: impact of treatment fidelity. *Contemporary Clinical Trials*, 37(2), 234–241.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11(3), 247–266.
- *Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, 45(2), 99–127.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315–340.
- *Musti-Rao, S., & Cartledge, G. (2007). Effects of a supplemental early reading intervention with at-risk urban learners. *Topics in Early Childhood Special Education*, 27(2), 70–85.

- National Association of School Psychologists. (2007). *Prevention and Intervention Research in the Schools [Position Statement]*. Bethesda, MD: Author.
- National Institutes of Health. (2011). *Learning disabilities research center request for application* (FOA Number RFA-HD-12-202). Washington, DC: Author. Retrieved from <http://grants.nih.gov/grants/guide/rfa-files/RFA-HD-12-202.html>
- National Research Council Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials, Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.
- *Nelson, J. R., Benner, G. J., & Gonzalez, J. (2005a). An investigation of the effects of a prereading intervention on the early literacy skills of children at risk of emotional disturbance and reading problems. *Journal of Emotional and Behavioral Disorders, 13*, 3–12.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39*(4), 374–396.
- *Nelson, J. R., Stage, S. A., Epstein, M. H., & Pierce, C. D. (2005b). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children, 72*(1), 29–45.
- *Niedringhaus, B. (2012). *Best practice in early reading intervention: Implementing a reading intervention program to reach below level readers* (Doctoral dissertation). Retrieved from PsycINFO. (2013-99230-264).
- No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).
- *Noltmeyer, A., Joseph, L., & Watson, M. (2014). Improving reading prosody and oral retell fluency: a comparison of three intervention approaches. *Reading Improvement, 51*(2), 221–232.
- *Nunnery, J. A., Ross, S. M., & McDonald, A. (2006). A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk, 11*(1), 1–18.
- *O'Connor, R. E., Bocian, K. M., Sanchez, V., & Beach, K. D. (2014). Access to a responsiveness to intervention model: Does beginning intervention in kindergarten matter?. *Journal of Learning Disabilities*.
- *O'Connor, R. E., Gutierrez, G., Teague, K., Checca, C., Kim, J. S., & Ho, T. H. (2013). Variations in practice reading aloud: Ten versus twenty minutes. *Scientific Studies of Reading, 17*(2), 134–162.
- *O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology, 87*(2), 202.
- *O'Connor, R. E., Notari-Syverson, A., & Vadasy, P. F. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. *Exceptional Children, 63*(1), 117–130.
- *O'Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology, 102*(1), 1–19.
- O'Donnell, C. L. (2007). *Fidelity of implementation to instructional strategies as a moderator of curriculum unit effectiveness in a large-scale middle school science quasi-experiment*. Dissertation Abstracts International, 68(08). (UMI No. AAT 3276564)
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84.
- *O'Shaughnessy, T. E., & Swanson, H. L. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities, 33*, 257–277.
- *Oudeans, M. K. (2003). Integration of letter-sound correspondences and phonological awareness skills of blending and segmenting: A pilot study examining the effects of instructional sequence on word reading for kindergarten children with low phonological awareness. *Learning Disability Quarterly, 26*(4), 258–280.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829.
- *Pericola Case, L., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., et al. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities, 43*, 402–417.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*(4), 477–492.
- Pressley, M., & Rankin, J. L. (1994). More about whole language methods of reading instruction for students at risk for early reading failure. *Learning Disabilities Research & Practice, 9*(3), 157–168.
- *Puhalla, E. M. (2011). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education, 32*(6), 471–481.

- *Pullen, P. C., Lane, H. B., Lloyd, J. W., Lloyd, J. W., Nowak, R., & Ryals, J. (2005). Effects of explicit instruction on decoding of struggling first grade students: A data-based case study. *Education and Treatment of Children*, 63–75.
- Quay, H. C. (1977). The three faces of evaluation: What can be expected to work. *Criminal Justice and Behavior*, 4, 341–354.
- *Reisener, C. D., Lancaster, A. L., McMullin, W. A., & Ho, T. (2014). A Preliminary Investigation of Evidence-Based Interventions to Increase Oral Reading Fluency in Children With Autism. *Journal of Applied School Psychology*, 30(1), 50–67.
- Roberts, G. (2016). Implementation Fidelity and Educational Science: An introduction. In G. Roberts, S. Vaughn, S. N. Beretvas, & V. Wong (Eds.), *Treatment Fidelity in Studies of Educational Intervention*. New York, NY: Routledge.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: : Introduction to the Special Series. *School Psychology Review*, 38(4), 445–553.
- Sanetti, L. M. H., Doherty, L. M., & Gallucci, J. (2013). Treatment integrity of interventions with children in School Psychology International from 1995–2010. *School Psychology International*, 0143034313476399.
- *Savage, R., Carless, S., & Stuart, M. (2003). The effects of rime-and phoneme-based teaching delivered by learning support assistants. *Journal of Research in Reading*, 26(3), 211–233.
- Scammacca, N., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4–12, 1914–2014. *Review of Educational Research*, 86(3), 756–800.
- Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). *Extensive reading interventions in grades K– 3: From research to practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin and Company.
- Sheridan, S. M., Swanger-Gagné, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review*, 38(4), 476.
- *Shippen, M. E., Reilly, A., & Dunn, C. (2008). The Effect of the Intensity of Spelling Instruction for Elementary Students At Risk for School Failure. *Journal of Direct Instruction*, 8(1), 19–28.
- *Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O. M., Simmons, L., Johnson, C., et al. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, 77(2), 207–228 Retrieved from <http://search.proquest.com/docview/847668046?accountid=14816>.
- *Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., et al. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities*, 40(4), 331–347 Retrieved from <http://search.proquest.com/docview/57243065?accountid=14816>.
- *Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., & Chaparro, E. (2015). An examination of the efficacy of a multitiered intervention on early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities*, 48(6), 602–621.
- Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, 47, 3–13 doi: 0022466911419516.
- *Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60(1), 40–56.
- *Ukrainetz, T. A., Ross, C. L., & Harm, H. M. (2009). An investigation of treatment scheduling for phonemic awareness with kindergartners who are at risk for reading difficulties. *Language, Speech, and Hearing Services in Schools*, 40(1), 86–100.
- U.S. Department of Education, Institute of Education Sciences (2011). Request for Applications: Education Research Grants (CFDA Number: 84.305A). Washington, DC.
- *Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly*, 20, 126–139.
- *Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, 33, 579–590.
- *Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002a). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research & Practice*, 17(4), 227–241.
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005a). Relative Effectiveness of Reading Practice or Word-Level Instruction in Supplemental Tutoring How Text Matters. *Journal of Learning Disabilities*, 38(4), 364–380.

- *Vadasy, P. F., Sanders, E. A., & Nelson, J. R. (2015). Effectiveness of supplemental kindergarten vocabulary instruction for English learners: A randomized study of immediate and longer-term effects of two approaches. *Journal of Research on Educational Effectiveness*, 8(4), 490–529.
- *Vadasy, P. F., & Sanders, E. A. (2008a). Code-oriented instruction for kindergarten students at risk for reading difficulties: a replication and comparison of instructional groupings. *Reading and Writing: An Interdisciplinary Journal*, 21, 929–963.
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98(3), 508.
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006c). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third-graders at risk for reading problems. *Remedial and Special Education*, 27, 363–378.
- *Vadasy, P. F., & Sanders, E. A. (2008b). Repeated reading intervention: outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology*, 100, 272–290.
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005b). Relative Effectiveness of Reading Practice or Word-Level Instruction in Supplemental Tutoring How Text Matters. *Journal of Learning Disabilities*, 38(4), 364–380.
- *Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading*, 13, 383–425.
- *Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, 102, 786–803.
- *Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: how language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15, 471–497.
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Code-oriented instruction for kindergarten students at risk for reading difficulties: a randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98, 508–528.
- *Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002b). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research. Practice*, 17(4), 227–241.
- *Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: beyond basic decoding skills. *Journal of Learning Disabilities*, 40, 508–525.
- *Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal*, 107(2), 153–180.
- Vaughn, S., & Swanson, E. A. (2015). Special education research advances knowledge in education. *Exceptional Children*, 82(1), 11–24.
- *Vernon-Feagans, L., Kainz, K., Ginsberg, M., Wood, T., & Bock, A. (2012). Targeted reading intervention: a coaching model to help classroom teachers with struggling readers. *Learning Disabilities Quarterly*, 35, 102–114.
- *Wang, C., & Algozzine, B. (2008). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education*, 22(4), 425–439.
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlín, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in Grades K-3. *Educational Psychology Review*, 3(28), 551–576.
- Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36, 541–561.
- *Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, 41, 126–142.
- *Watson, T. S., & Ray, K. P. (1997). The effects of different units of measurement on instructional decision making. *School Psychology Quarterly*, 12(1), 42.
- *Wehby, J. H., Falk, K. B., Barton-Arwood, S., Lane, K. L., & Cooley, C. (2003). The impact of comprehensive reading instruction on the academic and social behavior of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 11(4), 225–238.
- *Wehby, J. H., Lane, K. L., & Falk, K. B. (2005). An inclusive approach to improving early literacy skills of students with emotional and behavioral disorders. *Behavioral Disorders*, 155–169.
- What Works Clearinghouse. (2014). *Procedures and standards handbook* (version 3.0). Retrieved from <https://ies.ed.gov/ncee/wwc/>

- Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment Integrity A Review of Intervention Studies Conducted With Children With Autism. *Focus on Autism and Other Developmental Disabilities*, 21(1), 45–54.
- *Wolgemuth, J. R., Abrami, P. C., Helmer, J., Savage, R., Harper, H., & Lea, T. (2014). Examining the Impact of ABRACADABRA on Early Literacy in Northern Australia: An Implementation Fidelity Analysis. *The Journal of Educational Research*, 107(4), 299–311.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49(2), 156.
- Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation*, 33(4), 547–565.